



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2006

Human neural learning depends on reward prediction errors in the blocking paradigm

Tobler, Philippe N ; O'Doherty, J P ; Dolan, R J ; Schultz, W

Abstract: Learning occurs when an outcome deviates from expectation (prediction error). According to formal learning theory, the defining paradigm demonstrating the role of prediction errors in learning is the blocking test. Here, a novel stimulus is blocked from learning when it is associated with a fully predicted outcome, presumably because the occurrence of the outcome fails to produce a prediction error. We investigated the role of prediction errors in human reward-directed learning using a blocking paradigm and measured brain activation with functional magnetic resonance imaging. Participants showed blocking of behavioral learning with juice rewards as predicted by learning theory. The medial orbitofrontal cortex and the ventral putamen showed significantly lower responses to blocked, compared with nonblocked, reward-predicting stimuli. In reward-predicting control situations, deactivation in orbitofrontal cortex and ventral putamen occurred at the time of unpredicted reward omissions. Responses in discrete parts of orbitofrontal cortex correlated with the degree of behavioral learning during, and after, the learning phase. These data suggest that learning in primary reward structures in the human brain correlates with prediction errors in a manner that complies with principles of formal learning theory.

DOI: <https://doi.org/10.1152/jn.00762.2005>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-50277>

Journal Article

Published Version

Originally published at:

Tobler, Philippe N; O'Doherty, J P; Dolan, R J; Schultz, W (2006). Human neural learning depends on reward prediction errors in the blocking paradigm. *Journal of Neurophysiology*, 95(1):301-310.

DOI: <https://doi.org/10.1152/jn.00762.2005>

Human Neural Learning Depends on Reward Prediction Errors in the Blocking Paradigm

Philippe N. Tobler, John P. O'Doherty, Raymond J. Dolan and Wolfram Schultz
J Neurophysiol 95:301-310, 2006. First published 28 September 2005; doi:10.1152/jn.00762.2005

You might find this additional info useful...

This article cites 40 articles, 10 of which can be accessed free at:

<http://jn.physiology.org/content/95/1/301.full.html#ref-list-1>

This article has been cited by 15 other HighWire hosted articles, the first 5 are:

Expected value and prediction error abnormalities in depression and schizophrenia

Victoria B. Gradin, Poornima Kumar, Gordon Waiter, Trevor Ahearn, Catriona Stickle, Marteen Milders, Ian Reid, Jeremy Hall and J. Douglas Steele
Brain, June , 2011; 134 (6): 1751-1764.

[\[Abstract\]](#) [\[Full Text\]](#) [\[PDF\]](#)

Expected value and prediction error abnormalities in depression and schizophrenia

Victoria B. Gradin, Poornima Kumar, Gordon Waiter, Trevor Ahearn, Catriona Stickle, Marteen Milders, Ian Reid, Jeremy Hall and J. Douglas Steele
Brain, April 10, 2011; .

[\[Abstract\]](#) [\[Full Text\]](#) [\[PDF\]](#)

Striatal and Medial Temporal Lobe Functional Interactions during Visuomotor Associative Learning

Aaron T. Mattfeld and Craig E. L. Stark
Cereb. Cortex, March , 2011; 21 (3): 647-658.

[\[Abstract\]](#) [\[Full Text\]](#) [\[PDF\]](#)

Choice from Non-Choice: Predicting Consumer Preferences from Blood Oxygenation Level-Dependent Signals Obtained during Passive Viewing

Ifat Levy, Stephanie C. Lazzaro, Robb B. Rutledge and Paul W. Glimcher
J. Neurosci., January 5, 2011; 31 (1): 118-125.

[\[Abstract\]](#) [\[Full Text\]](#) [\[PDF\]](#)

Striatal and Medial Temporal Lobe Functional Interactions during Visuomotor Associative Learning

Aaron T. Mattfeld and Craig E. L. Stark
Cereb. Cortex, August 5, 2010; .

[\[Abstract\]](#) [\[Full Text\]](#) [\[PDF\]](#)

Updated information and services including high resolution figures, can be found at:

<http://jn.physiology.org/content/95/1/301.full.html>

Additional material and information about *Journal of Neurophysiology* can be found at:

<http://www.the-aps.org/publications/jn>

This information is current as of October 25, 2011.

Human Neural Learning Depends on Reward Prediction Errors in the Blocking Paradigm

Philippe N. Tobler,¹ John P. O'Doherty,² Raymond J. Dolan,² and Wolfram Schultz¹

¹Department of Anatomy, University of Cambridge, Cambridge; and ²Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, United Kingdom

Submitted 19 July 2005; accepted in final form 21 September 2005

Tobler, Philippe N., John P. O'Doherty, Raymond J. Dolan, and Wolfram Schultz. Human neural learning depends on reward prediction errors in the blocking paradigm. *J Neurophysiol* 95: 301–310, 2006. First published September 28, 2005; doi: 10.1152/jn.00762.2005. Learning occurs when an outcome deviates from expectation (prediction error). According to formal learning theory, the defining paradigm demonstrating the role of prediction errors in learning is the blocking test. Here, a novel stimulus is blocked from learning when it is associated with a fully predicted outcome, presumably because the occurrence of the outcome fails to produce a prediction error. We investigated the role of prediction errors in human reward-directed learning using a blocking paradigm and measured brain activation with functional magnetic resonance imaging. Participants showed blocking of behavioral learning with juice rewards as predicted by learning theory. The medial orbitofrontal cortex and the ventral putamen showed significantly lower responses to blocked, compared with nonblocked, reward-predicting stimuli. In reward-predicting control situations, deactivations in orbitofrontal cortex and ventral putamen occurred at the time of unpredicted reward omissions. Responses in discrete parts of orbitofrontal cortex correlated with the degree of behavioral learning during, and after, the learning phase. These data suggest that learning in primary reward structures in the human brain correlates with prediction errors in a manner that complies with principles of formal learning theory.

INTRODUCTION

Early classical theories proposed that reward-directed learning depends on the temporal contiguity between stimuli and reward (Pavlov 1927; Thorndike 1911). By contrast, in most modern learning theories (e.g., Mackintosh 1975; Pearce and Hall 1980; Rescorla and Wagner 1972), a discrepancy between actual and predicted reward (reward-prediction error) plays an important role for learning stimulus-reward associations. The Rescorla and Wagner model (1972) and its real-time extensions (temporal difference models) (Sutton and Barto 1981) postulate that learning is directly influenced by prediction errors that decrease gradually until the predictions match the outcome. According to other theories, prediction errors influence learning indirectly through changes of attention in subsequent trials (Mackintosh 1975; Pearce and Hall 1980).

The paradigmatic experiment to demonstrate the critical role for prediction errors is the blocking experiment (Kamin 1969). A typical blocking experiment generates differential prediction errors but maintains a similar amount of contiguity by rewarding a target stimulus that is presented in compound with a

pretrained stimulus. According to theory, as the pretrained stimulus fully predicts the reward, the reward fails to generate a substantial prediction error to the target stimulus. Behavioral analysis indicates that learning about the target stimulus in this situation is blocked despite its contiguity with reward.

Brain structures implicated in reward-directed learning include the orbitofrontal and temporal cortex, amygdala, striatum, insula, thalamus, and lateral hypothalamus (reviewed in Schultz 2000). The responses of midbrain dopamine neurons approximate that of a temporal difference signal (Montague et al. 1996; Schultz et al. 1997), and such a signal appears to be suitable for inducing synaptic modifications (Bao et al. 2001; Barto 1995; Brems et al. 2002; Reynolds et al. 2001; Wickens et al. 1996). These neurons show less activation to a blocked compared with a well-learned, reward-predicting, stimulus. This result can be explained by the induction of a positive prediction by the reward-predicting but not by the blocked stimulus. Omission of reward after a reward-predicting stimulus, but not after a blocked stimulus, depresses dopamine firing at the expected time of reward (Waelti et al. 2001). Depression of dopamine neurons reflects the negative prediction error induced by reward omission after the reward-predicting stimulus but not after the blocked stimulus. Blood-oxygen-level-derived (BOLD) activity in human ventral striatum and orbitofrontal cortex decreases in situations inducing negative prediction errors such as missed reward (Knutson et al. 2003), withheld reward (O'Doherty et al. 2003), and delayed reward (McClure et al. 2003). By contrast, situations inducing positive prediction errors elicit increases in BOLD signal in these dopamine-innervated areas on which we focused in the present study (McClure et al. 2003; O'Doherty et al. 2003, 2004).

Behavioral studies have established the role of prediction errors in human learning (for review, see De Houwer et al. 2001) by demonstrating blocking of aversive electrodermal conditioning (Hinchey et al. 1995), eyelid conditioning (Martin and Levey 1991), and causal learning (Dickinson 2001). However, it is unknown how the human brain processes reward-prediction errors during appetitive learning as tested in the blocking paradigm. The present study is based conceptually on the Rescorla-Wagner rule and its real-time extension, the temporal difference model, and tests the role of prediction errors in appetitive learning using the blocking paradigm by pairing abstract visual stimuli with fruit juice reward. We used functional magnetic resonance imaging (fMRI) to measure brain activations in prime reward structures during, and after,

Address for reprint requests and other correspondence: P. Tobler, Dept. of Anatomy, University of Cambridge, Cambridge CB2 3DY, UK (E-mail: pnt21@cam.ac.uk).

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

the learning phase and correlated the evoked activations with the degree of behavioral blocking.

METHODS

Subjects

Twenty-two right-handed healthy normal subjects (mean age: 27 yr; range: 19–50; 13 females) participated. Subjects were preassessed to exclude prior histories of neurological or psychiatric illness. They were asked to refrain from eating or drinking for 5 h prior to scanning and were thus in a mildly fluid-deprived state. Subjects rated their hunger, thirst, and the pleasantness of the juice (scale ranging from 0 = not at all hungry/thirsty to 10 = very hungry/thirsty or from –10 = very unpleasant to 10 = very pleasant). No specific action was taken to enforce subjects' compliance with dietary instructions, but ratings suggest compliance. Two subjects were psychology students, but none had knowledge of the blocking paradigm. All subjects gave informed consent, and the study was approved by the Joint Ethics Committee of the National Hospital for Neurology and Neurosurgery (UK).

Behavioral procedure

Subjects were placed on a moveable bed in the scanner, with light head restraint to limit head movement during image acquisition. Visual stimuli were presented for 3 s, and subjects viewed them through a mirror fitted on top of the head coil. Four abstract, complex visual stimuli, denoted as A, B, X, and Y, were used. Identities of the stimuli were counterbalanced across subjects. Stimulus A and stimulus compounds AX and BY were rewarded by fruit juice (20% dilution of commercial blackcurrant juice) at the end of the 3-s stimulus presentation, whereas stimulus B and the occasional presentations of stimuli X and Y went unrewarded. Intertrial intervals varied between 3 and 11 s according to a Poisson distribution with a mean of 6 s. Two 50-ml syringes contained the fruit juice and were attached to an SP220I electronic syringe pump (World Precision Instruments, Stevenage, UK). The pump was located in the scanner control room and delivered fixed quantities of 0.5 ml via a 6-m-long, 3-mm-diam polythene tube. The syringes were attached to a valve system. A stimulus presentation computer positioned in the control room controlled the apparatus. The same computer also received volume trigger pulses from the scanner. Both reward and picture delivery was controlled using Cogent 2000 software (Wellcome Department of Imaging Neuroscience, London, UK) as implemented in Matlab 6.0.

We employed a Pavlovian blocking procedure that comprised three consecutive phases during training and testing. In the first, pretraining phase (Fig. 1A), stimulus A was followed by liquid reward, marked with a "+" (A+), whereas stimulus B (B–) was not rewarded, denoted as "–". The two stimuli A+ and B– were presented in 10 trials, in random order, either on the left or the right side of a fixation cross. In each trial, the side of stimulus appearance was determined randomly. The task involved subjects indicating on which side of a central fixation cross the stimulus appeared by pressing one of two buttons on a button box. Subjects were positioned in the scanner during this training phase but not scanned. Scanning started with the second phase, in which stimulus X– appeared alongside A+ as a compound stimulus, followed by juice reward (AX+). Thus stimulus X– did not predict anything additional over and above to what stimulus A+ already predicted. Thus modern learning theory predicts that this stimulus would be blocked from learning. As a control, stimulus B– was shown simultaneously with stimulus Y–, and this compound was also rewarded (BY+). Theory predicts that stimulus Y– would not be blocked from learning, as the reward in BY+ trials was not predicted by any stimulus. Both AX+ and BY+ were presented in 15 trials. A+ and B– trials (10 trials) were also run in the second phase to maintain the previously learned associations. A+,

B–, AX+, and BY+ trials alternated randomly. In a subsequent third phase, stimuli X– and Y– were tested alone in 20 unrewarded trials that were randomly intermixed with A+, B– (20 trials), AX+, and BY+ (30 trials) trials.

Data acquisition and analysis

We had subjects rate the pleasantness of visual stimuli before and after the experiment on a scale ranging from 5 = very pleasant to (–5 = very unpleasant. Mean ratings were statistically evaluated by repeated-measures ANOVA. An interaction analysis between trial type and time (before and after the experiment) tested for changes in pleasantness ratings induced by the conditioning procedure. For linear regression analysis of brain activation data, the degree of behavioral blocking was determined as size of the difference [(pleasantness of Y– after experiment – pleasantness of Y– before experiment) – (pleasantness of X– after experiment – pleasantness of X– before experiment)]. In 15 subjects, this difference was positive, in agreement with a blocking effect, in 6 it was negative, and in 1, it was 0 (Fig. 1C for a separate analysis of subjects showing blocking and subjects not showing blocking).

We acquired gradient echo T2*-weighted echoplanar images (EPIs) with BOLD contrast on a Siemens Sonata 1.5 Tesla scanner (slices/volume, 40; repetition time, 3.6 s). 507 volumes were collected together with 5 "dummy" volumes at the start of the scanning session. Scan onset times varied relative to stimulus onset times. A T1-weighted structural image was also acquired for each subject. Signal dropout in basal frontal and medial temporal structures due to susceptibility artifact was reduced by using a tilted plane of acquisition (30° to the anterior commissure-posterior commissure line, rostral > caudal). Imaging parameters were: echo time, 50 ms; field-of-view, 192 mm; in-plane resolution, 3 mm; slice thickness, 2 mm; interslice gap, 1 mm. High-resolution T1-weighted structural scans were coregistered to their mean EPIs and averaged together to permit anatomical localization of the functional activations at the group level.

Statistical Parametric Mapping (SPM2) served to spatially realign functional data, normalize them to a standard EPI template, and smooth them using a Gaussian kernel with a full width at half-maximum of 10 mm. Functional data were then analyzed by constructing a set of 3-s stick functions at the event-onset times for each of the six trial types (A+, B–, AX+, BY+, X–, and Y–), corresponding to the duration of visual stimulus presentation. We used a standard rapid event-related fMRI approach in which evoked hemodynamic responses to each trial type are estimated separately by convolving a canonical hemodynamic response function with the onsets for each trial type and regressing these trial regressors against the measured fMRI signal (Dale and Buckner 1997; Josephs and Henson 1999). This approach makes use of the fact that the hemodynamic response function summates in an approximately linear fashion over time (Boynton et al. 1996). By presenting trials in random order and using variable intertrial intervals, it is possible to separate out fMRI responses to rapidly presented events without waiting for the hemodynamic response to reach baseline after each single trial (Dale and Buckner 1997; Josephs and Henson 1999).

Subject-specific movement parameters were modeled as covariates of no interest. Trial type-specific estimates of neural activity (betas), corresponding to the height of the HRF, were computed independently at each voxel for each subject, using the general linear model (GLM) (see Friston et al. 1994 for detailed description of how the GLM is used in an imaging context). The estimated GLM parameter beta summarized the amount of variance in each fMRI time series accounted for by the events in the experiment. More specifically, the GLM conforms to $Y = \beta X + \epsilon$, where β (parameter estimate) reflects the strength of covariance between Y (the data) and X (canonical response function for a given condition such as A+ or B–), given error ϵ . Parameter estimates were contrasted against each other to assess differential model fit for different conditions. Using random-

effects analysis, these contrasts were entered into a series of one-way t-test, simple regressions or repeated-measures ANOVAs with non-sphericity correction where appropriate. MarsBaR (Brett et al. 2002) served to compute mean activations in two functional regions of interest (10 mm sphere around peak voxels in the right ventral putamen; 27/−9/−9, 26/6/−8) described previously (O'Doherty et al. 2003, 2004). For time course plots, we also used MarsBaR (Brett et al. 2002), making no assumptions about the shape of activations, and applying eight finite impulse responses per trial, each response separated from the next by one scan (3.6s). The dependent measure in time course plots is percentage signal change measured within spheres of 10 mm around peak voxels.

Model setup and contrasts

The data were analyzed using two different approaches. In one analysis, a temporal difference model was used as previously described (O'Doherty et al. 2003; Schultz et al. 1997) to analyze learning in AX+ and BY+ trials. Briefly, the temporal difference model suggests that prediction errors are computed according to $\delta(t) = r(t) + \gamma(t+1) - V(t)$ where $V(t)$ corresponds to the predicted value V at time t in the trial, $r(t)$ corresponds to the reward at time t , and γ corresponds to a factor for discounting rewards which occur later in time. Thus the temporal difference model suggests that prediction errors correspond to the difference between predicted values at consecutive time steps. At the end of each trial, these are used to update the values of all the stimuli present in that trial. For example, in initial BY+ trials, the value of B− is low, but the reward occurs, and value is attributed mostly to Y−. After learning, A+ and Y− elicit a positive prediction, whereas B− and X− do not. Responses to stimuli A+ and B− were modeled as phasic increases at the time of conditioned stimuli; responses to Y− and X− were modeled as phasic increases at the time of conditioned stimuli and phasic decreases at the usual time of unconditioned stimuli. We tested for regions showing an activation pattern that fitted the model better for A+ than B− or Y− than X−. Thus the effect of reward prediction versus prediction of no reward was examined in the contrast of (A+) − (B−), and the effect of a nonblocked stimulus versus a blocked stimulus was examined in the contrast of (Y−) − (X−). The conjoint effect of these two contrasts was examined in a conjunction of (A+) − (B−) and (Y−) − (X−), a conjunction that tests for responses that are selective for reward-predicting stimuli and are more activated by a nonblocked than a blocked stimulus. Bar plots show contrast estimates corresponding to the average fit of the effects of interest with the model.

In a second analysis, the effect of learning the associative strength of a novel reward-predicting stimulus was examined in the contrast of (AX+) − (BY+), after convolving the regressors of AX+ and BY+ with an exponential function that had a half-life equal to 1/4 of the session length. This exponential function models asymptotical acquisition of associative strength in BY+ trials but not in AX+, similar to how learning theories capture the negatively accelerated increase of associative strength between conditioned and unconditioned stimulus during learning. The effect of gradual reduction in prediction errors was examined in the opposite contrast, (BY+) − (AX+), both convolved with the exponential function. Thresholding strategy has been described previously (O'Doherty et al. 2002–2004). For each analysis, in a priori brain regions identified in previous neuroimaging studies of appetitive conditioning (O'Doherty et al. 2002, 2003), including ventral striatum and orbitofrontal cortex, we report activations surviving a threshold of $P < 0.001$ uncorrected. Reported voxels conform to Montreal Neurological Institute (MNI) coordinate space. For display, the right side of the image corresponds to the right side of the brain, and functional activations at $P < 0.001$ are overlaid on the average structure of participating subjects.

RESULTS

Behavioral paradigm

The blocking paradigm employed four visual stimuli leading to different levels of learning. Stimulus A+ was followed by the delivery of juice, whereas control stimulus B− was not followed by reward (Fig. 1A). After learning, the reward following stimulus A+ and the absence of reward following stimulus B− were fully predicted and should not generate prediction errors. During subsequent compound training, two stimuli, X− and Y−, were presented simultaneously with A+ and B−, respectively, and both the AX+ and BY+ compounds were paired with reward. In AX+ trials, the reward was already fully predicted by the pretrained stimulus A+, and therefore should not have generated a prediction error. Conversely, in the BY+ control trials, the reward was predicted by neither stimulus, and the occurrence of reward should have generated a prediction error. The critical test involved presentation of stimulus X− and stimulus Y− alone. Stimulus X− was paired with reward in the absence of a prediction error and, according to theory, should not have been learned (blocking). Conversely, control stimulus Y− was paired with reward in the presence of a prediction error and should have been learned as an effective reward predictor.

Behavior

Subjects rated the pleasantness of visual stimuli before and after the learning experiment. There were no significant differences in pleasantness rating before learning for the comparisons between stimuli A+ and B− and between stimuli X− and Y− [for all analyses, $F(1,21) < 1.85$, $P > 0.18$]. However, in both cases, trial type interacted with time (before vs. after learning), indicating that the pleasantness of the visual stimuli changed during conditioning [A+ vs. B−, $F(1,21) = 5.91$; X− vs. Y−, $F(1,21) = 4.50$, both $P < 0.05$]. Inspection of the data revealed that the learning procedure had increased the pleasantness of stimuli A+ and Y− but not of stimuli B− and X− (Fig. 1B). After learning, the pleasantness of stimulus A+ was not significantly different from that of stimulus Y− and the pleasantness of stimulus B− was not significantly different from that of X− ($P > 0.34$). These results suggest that stimulus A+ had been learned as a valid reward predictor, whereas stimulus B− did not predict reward, and appetitive learning was blocked for stimulus X− but not for stimulus Y−.

Inspection of individual pleasantness ratings indicated that 15 subjects showed changes compatible with a blocking effect: pleasantness of stimulus Y− increased, whereas that of stimulus X− did not. Conversely, six subjects showed decreases of pleasantness for stimulus Y− but not for stimulus X− (Fig. 1C), and one subject showed no changes for either X− or Y−. Could the differential increase in pleasantness of Y− and X− have been due to factors other than the experimental manipulation? We found no correlation between the individual degree of blocking and contingency awareness, age, hunger, thirst, juice pleasantness, and scan-to-scan movements (for all correlations, $|r| < 0.32$ and $P > 0.18$).

To investigate blocking with an additional behavioral measure, we recorded reaction times in 12 participants that showed blocking in the pleasantness ratings. Reaction times showed an

overall difference between trial types [ANOVA, $F(5,2351) = 3.85$, $P < 0.05$]. Subjects responded more quickly with reward-predicting stimulus A+ than with neutral stimulus B- [737.2 ± 10.9 (SE) ms vs. 778.4 ± 13.7 ms; $P < 0.05$] and with reward-predicting stimulus Y- than with blocked stimulus X- (729.6 ± 13.0 vs. 753.0 ± 16.1 ms; $P < 0.05$). There were no significant reaction time differences between AX+ and BY+ trials (745.2 ± 11.0 vs. 738.3 ± 9.5 ms; $P > 0.5$). These results suggest appetitive learning of reward-predicting stimulus A+ but not of neutral stimulus B- and blocking of appetitive learning for stimulus X- compared with reward-predicting stimulus Y-.

Putamen activation reflecting blocking and reward expectation

We tested differential blocking of learning by modeling neural responses to control stimulus Y- and blocked stimulus

X- with a phasic positive response at the time of the stimulus and a negative prediction-error response at the time of the omitted reward. We performed a region of interest (ROI) analysis in 15 subjects showing behavioral blocking by measuring the activation in a 10-mm sphere centered on two previously reported peaks of reward-prediction-error responses in the ventral putamen (O'Doherty et al. 2003, 2004). Activations were stronger for Y- compared with X- (paired t -test, both $P < 0.05$, small volume correction; Fig. 2A) and failed to correlate with movement parameters ($|r|$ for all parameters < 0.53 and P for all > 0.12). In a conjunction analysis, we found that the right ventral putamen ($27/3/-6$; $z = 3.61$) was more activated by control stimulus Y- than by blocked stimulus X- and likewise more by reward-predicting stimulus A+ than by neutral stimulus B- (Fig. 2C; Table 1, top, for additional activations). These data suggest that activation in the ventral putamen was blocked together with behavioral learning in the absence of a reward-prediction error.

The differential blocking is also observed in the time courses of putamen activation in these ROIs. Reward-predicting stimulus Y- elicited a greater increase in brain activation than blocked stimulus X- (Fig. 2B). The activation to stimulus Y- was followed by a deactivation, reflecting the negative prediction error induced by reward omission after a reward-predicting stimulus. In addition, we investigated activations related to the expectation of reward in A+ compared with B- trials. Compared with the expectation of no reward, both left ($-18/3/-3$; $z = 4.03$) and right ($18/-3/-6$; $z = 3.63$) putamen were activated when subjects expected a reward (Fig. 3A). The activation extended into the globus pallidus. Time course analysis of the phasic activations in the left and right ventral putamen confirmed the differential responses to stimulus A+ and B- (Fig. 3B; see Table 2, top, for additional activations).

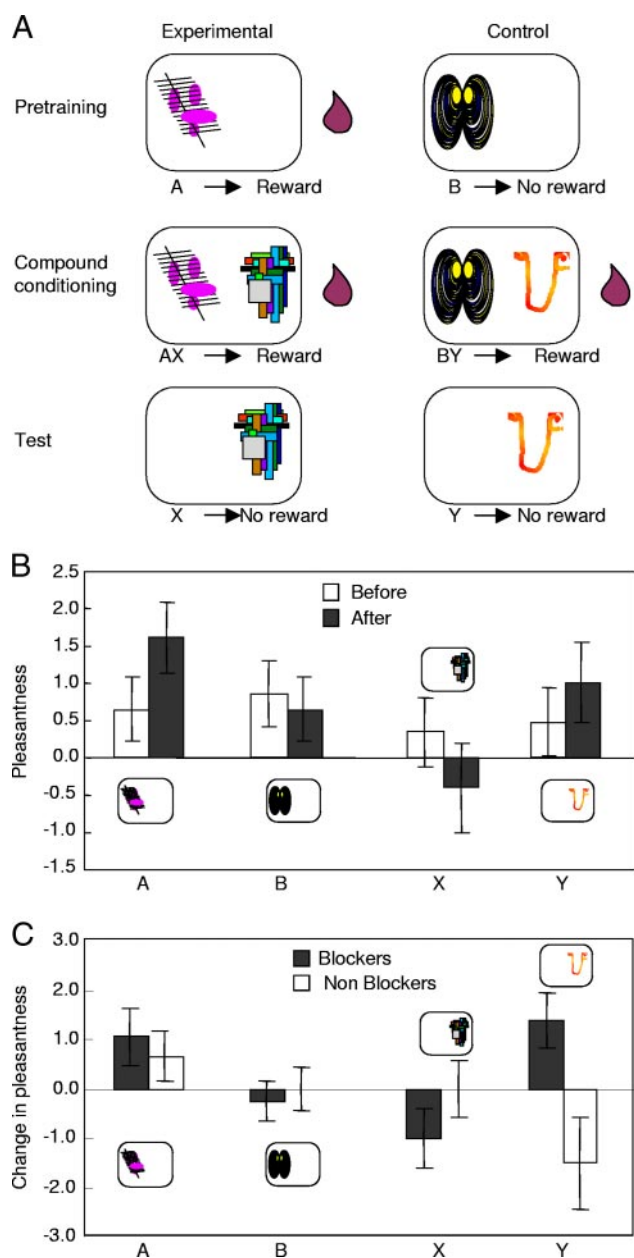


FIG. 1. Experimental design and pleasantness ratings in the Pavlovian blocking paradigm. **A**: experimental design. During pretraining, reward-predicting stimulus A+ but not neutral stimulus B- was followed by a drop of juice reward (letters denote visual stimuli presented on a computer monitor, + denotes reward, - denotes no reward). During compound conditioning, 2 additional stimuli, X- and Y-, appeared together with A+ and B- in rewarded compounds AX+ and BY+, respectively. In AX+ trials, reward was already predicted by stimulus A+, and no prediction error should have occurred. Stimulus X- should therefore have been blocked from learning. Conversely, in BY+ trials, reward was not predicted by stimulus B-, and a prediction error should have occurred. Stimulus Y- should therefore have been learned as a reward-predicting stimulus. During unrewarded test trials, stimuli X- and Y- were presented alone. In each trial, subjects indicated whether stimuli appeared on the left, right, or both sides of the screen. The position of stimuli on the screen varied randomly. During compound conditioning, A+, B-, AX+, and BY+ alternated randomly. During the test phase, all 6 trial types alternated randomly. The compound conditioning phase was immediately followed by the test phase. **B**: average absolute pleasantness ratings of all subjects. Compared with the beginning of the experiment, pleasantness ratings were increased at the end of the experiment for reward-predicting stimulus A+ and prediction error-eliciting control stimulus Y- but not for neutral stimulus B- and blocked stimulus X-. Pleasantness ratings were taken on a dimensionless scale of -5 (very unpleasant) to +5 (very pleasant). **C**: average change in pleasantness rating in subjects showing blocking compared with subjects not showing blocking. Over the course of the experiment, pleasantness of A+ but not B- and X- increased in both subjects showing blocking behaviorally ($n = 15$) and in subjects not showing blocking ($n = 6$). Pleasantness of Y- increased in subjects showing blocking but decreased, for unknown reasons, in subjects not showing blocking.

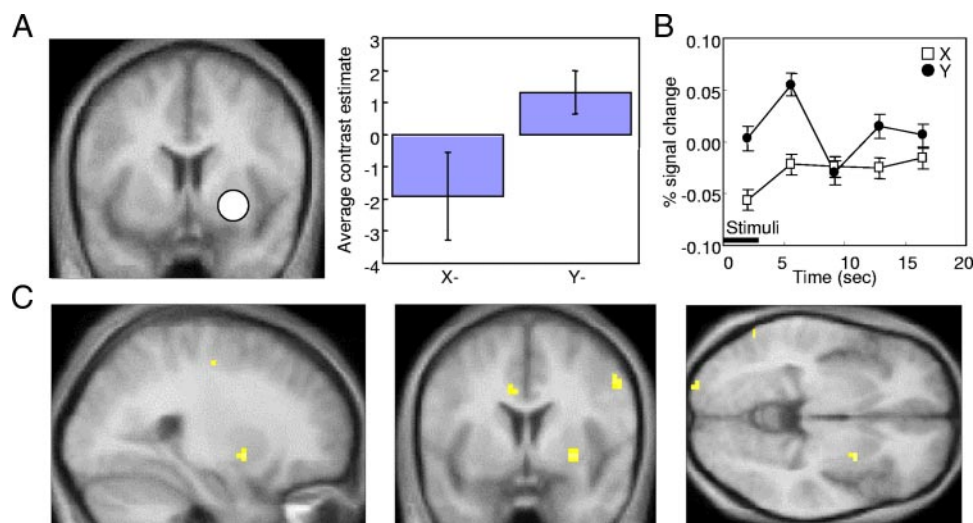


FIG. 2. Differential reward-prediction-related activity at the time of stimuli Y- and X- and of omitted reward in the ventral putamen during the blocking experiment. *A*: region of interest analysis. *Left*: 10-mm sphere centered on a striatal voxel shown previously to report prediction errors (O'Doherty et al. 2004). *Right*: bar plots show contrast estimates (dimensionless) corresponding to the average fit for X- and Y- when modeled with an activation at the time of conditioned stimuli and a deactivation at the usual time of reward within region of interest shown *left*. *B*: event-related time courses to X- and Y-, averaged within the region of interest shown in *A* and across subjects. Stimuli occurred at time = 0, were displayed for 3 s, and not followed by reward. Intertrial intervals varied between 3 and 11 s according to a Poisson distribution with a mean of 6 s. All 6 trial types alternated randomly, and scan onsets occurred randomly with regard to stimulus onsets. In this and all following time course plots, the hemodynamic response is plotted as percentage signal change and occurs with the usual lag of ~5 s. Error bars correspond to SE. To characterize the shape of the responses in time course plots, regressors have been convolved with finite impulse responses rather than hemodynamic response functions. The bin size corresponded to the scan repetition time (3.6 s). *C*: conjunction of reward-predicting stimulus A+ vs. neutral stimulus B- and reward-predicting stimulus Y- vs. blocked stimulus X-. This analysis was restricted to subjects showing blocking behaviorally.

Decrease of neural prediction error in ventral striatum during learning

During learning, a gradual (asymptotic) decrease of prediction error occurs at the time of the gradually better predicted reward (Rescorla and Wagner 1972; Sutton and Barto 1981). We specifically investigated whether brain activations would show better fits with asymptotic decreases in BY+ compared with AX+ trials as differential learning progressed. We found that in the 15 subjects showing blocking behaviorally, activation in the ventral striatum fitted better for BY+ than AX+ trials with an asymptotically decreasing learning function, corresponding to gradually reduced prediction error responses (Fig. 4; $-15/-3/-12$; $z = 3.98$).

TABLE 1. Activated areas in the conjunction of A+ versus B- and Y- versus X- and in the single contrast of Y- versus X-, restricted to subjects showing blocking behaviorally

A+ versus B- and Y- versus X-	Hemisphere	x	y	z	z Score
Striatum (ventral putamen)	Right	27	3	-6	3.6
Cingulate (posterior)	Left	-9	-30	45	3.7
Cingulate (middle)	Left	-6	3	33	3.2
Cerebellum	Left	-27	-60	-45	3.3
	Left	-30	-57	-42	3.2
Lateral frontal cortex	Right	54	0	39	3.7
	Right	18	-6	63	3.3
Lateral prefrontal cortex	Right	54	27	27	3.5
Y- versus X-					
Orbitofrontal cortex	Left	-18	36	-6	3.9
Cingulate (posterior)	Left	-9	-36	36	3.1

Correlation between behavioral and orbitofrontal responses to stimuli blocked from learning

We performed a linear regression analysis of differential brain activation following reward-predicting stimulus Y- compared with blocked stimulus X- against the individual degree of behavioral blocking. All subjects were included in the analysis, irrespective of their blocking behavior. We found a significant correlation in the medial orbitofrontal cortex (Fig. 5A; peak at $-18/30/-6$; $z = 3.27$). This region overlapped with the orbitofrontal region that showed stronger activation for Y- than X- in the previous analysis restricted to subjects with behavioral blocking (peak at $-18/36/-6$; $z = 3.89$; Table 1, bottom).

Further relationships between behavioral and learning related neural responses are revealed by the time courses of activation in a 10-mm orbitofrontal sphere (centered around $-18/36/-6$). The initial activations differed between stimuli Y- and X- only in subjects that showed behavioral blocking with stimulus X- being ineffective in these subjects (Fig. 5, B and C). Furthermore, we also used an interaction analysis to compare activations in brain regions between subjects showing behavioral blocking and no blocking. We found greater activations for Y- compared with X- in the medial orbitofrontal cortex (peak at $-18/33/-9$; $z = 3.74$; Fig. 5D) and posterior cingulate ($-9/-39/39$; $z = 3.71$) only in subjects showing blocking, whereas no such differences occurred in subjects that failed to show blocking.

Correlation between behavioral blocking and asymptotic orbitofrontal response increases during learning

During learning, a gradual (asymptotic) increase of associative strength of the conditioned stimulus occurs simultaneously

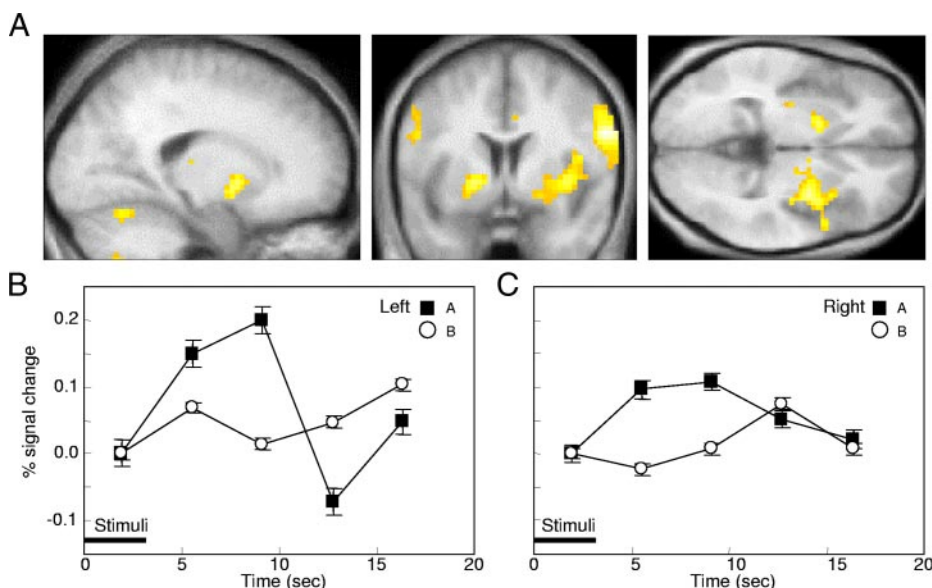


FIG. 3. Activation in putamen by expectation of reward. A: compared with neutral stimulus B–, reward-predicting stimulus A+ activated the putamen bilaterally. Activations extended into the globus pallidus and the right insula. The cortical and cerebellar activations were not expected a priori and did not survive whole brain correction. B and C: event-related time courses of responses to stimuli A+ and B– averaged across subjects within the left (–18/3/–3; B) and right (18/3/–6; C) ventral putamen. Stimuli occurred at time = 0 and were displayed for 3 s. A+ but not B– was followed by reward at the end of the 3-s stimulus duration. This analysis was performed over all subjects, irrespective of behavioral blocking.

with decreases in prediction errors. We specifically searched for activations showing better fit with a gradually increasing asymptotic learning function in BY+ trials compared with AX+ trials during progressive differential learning and correlated the obtained differential increases with the degree of individual behavioral blocking. The activity in an anterior region of orbitofrontal cortex correlated with the degree of behavioral blocking across all subjects during learning in BY+ trials (Fig. 6; –27/36/–15; $z = 3.49$). Thus responses in orbitofrontal cortex increased asymptotically during learning, and these increases reflected the degree to which subjects showed behavioral learning in the blocking paradigm.

DISCUSSION

These data suggest that human brain structures acquire responses to reward-predicting stimuli proportional to the degree of reward prediction error as suggested by formal learning theory. Activation of the ventral putamen was sensitive to blocking: in subjects showing blocking behaviorally, stimuli that were learned elicited prediction-error responses in the ventral putamen that were greater than those elicited by stimuli presented contiguously with reward and that failed to induce prediction errors. Distinct regions in the orbitofrontal cortex correlated with the degree to which subjects showed the blocking effect and showed asymptotic increases during learning the

associative strength of a stimulus predictive of reward. The results are compatible with basic concepts of learning theory and indicate a correlation of activation in human basal ganglia and orbitofrontal regions with learning induced by prediction errors in the context of a formal blocking paradigm.

In the present experiment, humans rated reward-predicting stimuli as more pleasant than neutral and blocked stimuli (evaluative conditioning) (for review, see De Houwer et al. 2001). The results confirm that human appetitive learning can be blocked, presumably due to the lack of prediction error caused by a previously established prediction of reward. It thus appears that appetitive learning is governed by similar associative mechanisms as other forms of Pavlovian conditioning such as aversive electrodermal and eyelid conditioning (Hinchey et al. 1995; Martin and Levey 1991). Apparently the mere contiguity between a stimulus and reward is insufficient for an increase in pleasantness of that stimulus. Rather, learning depends crucially on the presence of an error in the prediction of an appetitive outcome.

The pleasantness ratings for stimulus X– decreased over the course of the experiment. However, absolute differences in stimulus ratings over the experiment should not be interpreted without additional control stimuli that were never paired with reward and that were not included in the present experimental design. Irrespective of this result, the behavioral ratings suggest that the relative differences in pleasantness ratings between X– and Y– changed in the direction compatible with a blocking effect in 75% of subjects. Our study also found that thirst, hunger, juice pleasantness, age, and contingency awareness did not correlate significantly with the degree of the blocking effect. Possible explanations for the partial effectiveness of our experimental parameters include insufficient reward intensity and a high ratio of stimulus-reward interval to intertrial interval.

During learning, prediction errors gradually decrease, and the associative strength (motivational value) of stimuli increases. In the present experiment, the associative strength of the control stimulus, Y–, gradually increased while subjects learned about the predictive relation between Y– and reward in BY+ trials. Rostral orbitofrontal activations increased

TABLE 2. Areas activated by A+ versus B– over all subjects

A+ versus B–	Hemisphere	x	y	z	z Score
Striatum (ventral putamen)	Left	–18	3	–3	4.0
	Right	18	–3	–6	3.6
	Right	9	–3	–3	3.4
Cingulate (middle)	Right	6	3	36	3.2
	Right	9	18	39	3.2
Cerebellum	Left	–15	–66	–24	3.8
	Right	18	–66	–24	4.1
	Left	–6	–42	–33	3.9
Lateral frontal cortex	Left	–51	6	33	3.5
	Right	63	6	27	4.6
Insula	Right	45	3	12	3.5

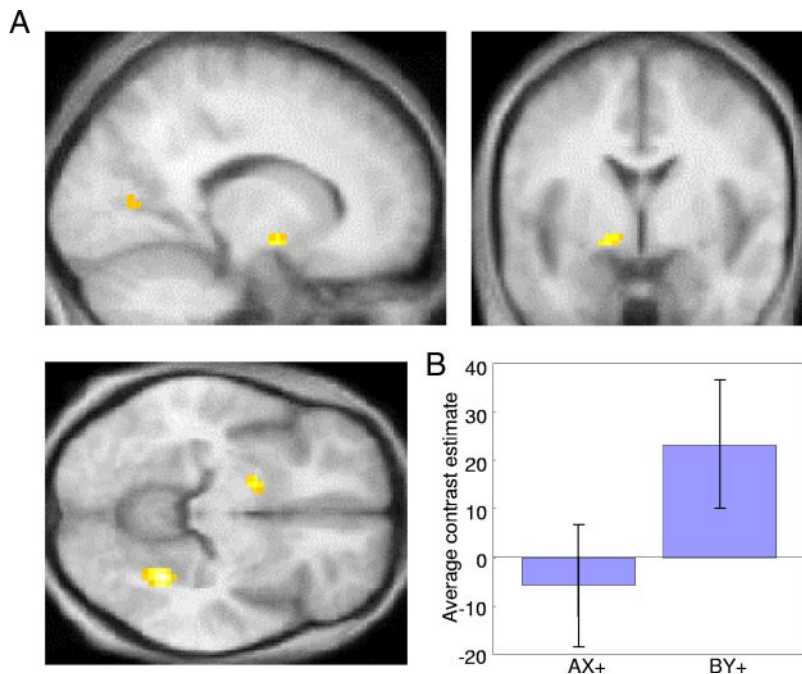


FIG. 4. Differential activations at the time of the reward in ventral striatum during reduction of prediction errors during learning. *A*: regions showing significantly better fits ($P < 0.001$) with modeled asymptotic decreases in activation during conditioning in BY+ compared with AX+ trials. Thus prediction-error related striatal activation at the time of reward decreased more in BY+ compared with AX+ trials during learning. *B*: bar plots showing contrast estimates (dimensionless) corresponding to the average fit of AX+ and BY+ with an asymptotically decreasing learning function. Error bars correspond to 95% confidence intervals. This analysis was performed only in subjects that showed blocking behaviorally.

asymptotically during learning as a function of the degree of behavioral blocking. The asymptotical increase in orbitofrontal activation is compatible with the acquisition of associative strength during learning proposed by learning theories (Rescorla and Wagner 1972; Sutton and Barto 1981). Neurophysiological studies reported reward expectation and cue-related activity of orbitofrontal neurons that changed together with behavioral indicators of learning (Schoenbaum et al. 2003; Tremblay and Schultz 2000b), although the relation to learning theories was less well explored in these studies. The present data extend these neurophysiological studies by suggesting that

the human orbitofrontal cortex processes the acquisition of associative strength of conditioned stimuli during learning according to a formal learning curve.

The omission of a predicted reward reflects an outcome that is worse than expected, and learning theory suggests that it elicits a negative prediction error. In the present study, control stimulus Y- predicted reward as it was paired with reward and prediction error in BY+ trials. Thus when stimulus Y- was presented in unrewarded test trials, a positive prediction should have occurred at the time of the stimulus (reward prediction) and a negative prediction error should have occurred at the

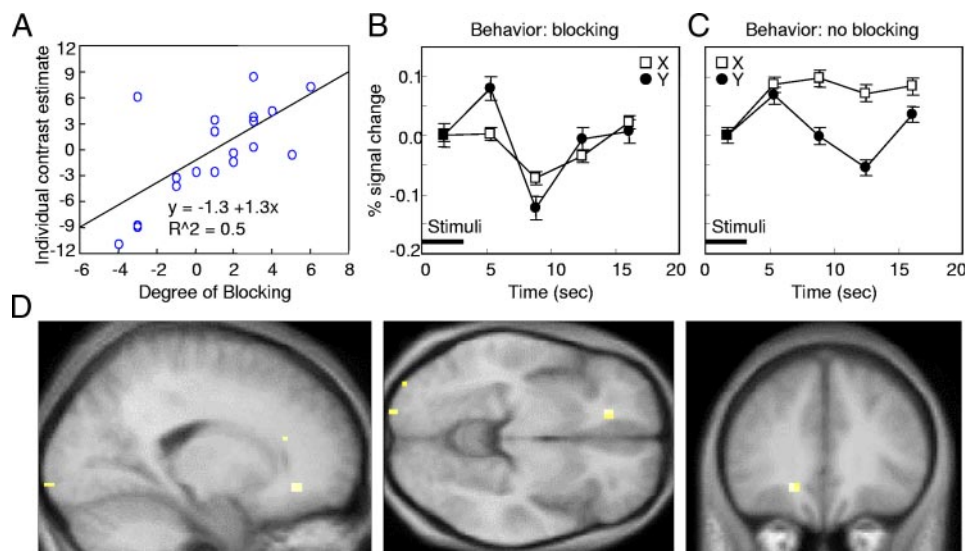


FIG. 5. Relationship of orbitofrontal activation to degree of behavioral blocking. *A*: correlation with degree of blocking. Across all subjects, the differential activation to Y- compared with X-, plotted as dimensionless contrast estimates, in the medial orbitofrontal cortex was higher with larger pleasantness differences between nonblocked stimulus Y- and blocked stimulus X-. The individual data points are from individual subjects. *B* and *C*: average event-related time courses in medial orbitofrontal cortex (10 mm sphere around -18/36/-6) to stimuli Y- and X- in subjects showing blocking behaviorally (*B*) and in subjects not showing blocking behaviorally (*C*). Stimuli occurred at time = 0, were displayed for 3 s, and not followed by reward. *D*: comparison of subjects showing blocking behaviorally with subjects not showing blocking revealed blocked acquisition of activation in medial orbitofrontal cortex in subjects that showed behavioral blocking but not in subjects that failed to show blocking. Both activation and deactivation for Y- was stronger than for X- in the blocking group. Conversely, the nonblocking group showed similar activation to Y- and X-.

A

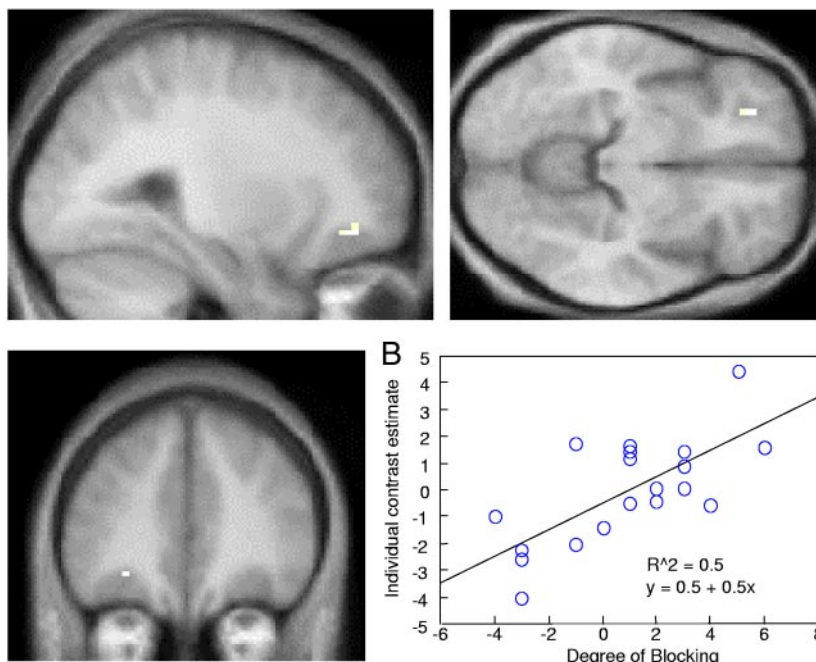


FIG. 6. Relationship between asymptotic increases of orbitofrontal activation and degree of behavioral blocking during acquisition of associative strength. *A*: regions showing a correlation between differential fit in BY+ and AX+ trials, both modeled with gradual increases, and degree of behavioral blocking. Thus orbitofrontal activation increased more in BY+ compared with AX+ trials the more subjects showed blocking behaviorally ($P < 0.001$; $r = 0.7$). *B*: scatter plot of differential activation increases during BY+ compared with AX+ trials, displayed in terms of dimensionless contrast estimates, and degree of blocking. Across all subjects, degree of blocking was quantified as pleasantness differences between non-blocked stimulus Y- and blocked stimulus X-. The individual data points are from individual subjects. This analysis was performed over all subjects, irrespective of whether or not they showed blocking behaviorally.

usual time of reward (reward omission). In Y- trials, ventral putamen and orbitofrontal cortex showed activations at the time of conditioned stimuli followed by deactivations at the time of reward (Figs. 2*B* and 5*B*). Thus both the ventral putamen and the orbitofrontal cortex appeared to code a bidirectional prediction error signal with increased activation induced by positive prediction errors and decreased activation induced by negative prediction errors. In lateral regions of the prefrontal cortex, both positive and negative prediction errors elicit increased activation (Fletcher et al. 2001; O'Doherty et al. 2003). Such a unidirectional prediction error signal would be reminiscent of the one proposed by attentional theories (Mackintosh 1975; Pearce and Hall 1980). Thus different regions may process different prediction error signals, and striatal and orbitofrontal regions appear to code a bidirectional signal.

Activations in the ventral putamen appeared to be sensitive to prediction errors in being stronger to nonblocked control than to blocked test stimuli, and they decreased during learning in subjects showing blocking behaviorally. Simple contiguity pairing of a stimulus with reward, as in the case of the blocked stimulus, was insufficient to activate the ventral putamen. Rather a prediction error, as elicited by the nonblocked control stimulus, was necessary for putamen activation. Results from previous imaging studies suggest that activation of the ventral putamen reflects reward-prediction errors (McClure et al. 2003; O'Doherty et al. 2003). The present study extends these findings by showing that the ventral putamen processes prediction errors in the blocking paradigm that tests for the crucial role of such prediction errors in learning stimulus-reward associations. Thus rewards that produce prediction errors correlate with putamen activations and behavioral learning, and the activation of the ventral putamen at the time of the reward may reflect a teaching signal as proposed by current learning theories and their real-time extensions (Rescorla and Wagner 1972; Sutton and Barto 1981).

The present results suggest that the degree of medial orbitofrontal activation by the nonblocked control stimulus compared with the blocked experimental stimulus correlated with the degree of behavioral blocking across all subjects. Correspondingly, medial orbitofrontal cortex is more activated by the nonblocked stimulus than by the blocked stimulus in subjects showing behavioral blocking but not in subjects without blocking. Single-cell recordings indicate that some orbitofrontal neurons respond to unpredicted reward delivered outside the task (Tremblay and Schultz 2000a) and to omitted reward when the animal makes an error (Thorpe et al. 1983). Results from a previous functional imaging experiment show that the orbitofrontal cortex is activated by unexpected rewards and depressed by unexpected reward omissions, indicating the explicit processing of reward prediction errors (O'Doherty et al. 2003). The present results extend these findings by suggesting that activations in orbitofrontal cortex may follow the systematic experimental manipulations of prediction errors to the degree to which individual subjects follow them behaviorally. Taken together the orbitofrontal cortex appears to process errors in reward prediction according to formal assumptions of learning theory.

The presently observed activations in the ventral putamen and the orbitofrontal cortex resemble the stronger responses of dopamine neurons for reward-predicting stimuli compared with neutral stimuli (Ljungberg et al. 1992; Waelti et al. 2001). Furthermore, dopamine neurons acquire weaker responses to stimuli that are blocked from learning compared with control stimuli that are being learned in the presence of a reward-prediction error (Waelti et al. 2001). These similarities suggest that learning theories can account for both phasic dopamine firing and activation of ventral putamen and orbitofrontal cortex. Thus dopamine, orbitofrontal and striatal regions appear to signal prediction errors and acquire responses to conditioned stimuli dependent on prediction errors.

Both putamen and orbitofrontal cortex regions are innervated by dopamine neurons (Groves et al. 1994; Lynd-Balta and Haber 1994; Williams and Goldman-Rakic 1998). Given that the hemodynamic responses measured by fMRI may reflect mainly inputs to an activated region rather than the spiking activity of projection neurons (Logothetis et al. 2001), it is tempting to suggest that the prediction-error-dependent learning observed presently might be driven by dopamine inputs. Alternatively, dopamine might influence different neuronal processes in the two target structures. For example, reward-processing neurons in the orbitofrontal cortex might be preferentially involved in detection, perception, and expectation of reward, whereas those in the striatum might also incorporate reward information into motor preparation (Pasupathy and Miller 2005; Schultz 2000). Dopamine might also affect blood flow through dilatory effects on the vascular system (Amenta et al. 2000; Hughes et al. 1986), and this effect could potentially contribute to the present activations. However, it is not clear what the time scale of such an effect would be and whether this would contribute to rapid event-related (phasic) activations of the type seen here.

Based on previous results, our hypotheses were primarily restricted to the striatum and orbitofrontal cortex. However, prediction error coding may be operational in several other brain structures as well. For instance, cingulate, cerebellum, superior colliculus, frontal, parietal and occipital cortex, locus coeruleus, and nucleus basalis show various forms of prediction error processing (for review, see Schultz and Dickinson 2000). Some of these regions showed activation in the present study in situations eliciting prediction errors. For example the posterior cingulate was more activated by control stimulus Y than by blocked stimulus X, and the activations were related to the degree to which subjects showed blocking behaviorally. Posterior cingulate neurons respond to the unexpected delivery and omission of reward (McCoy et al. 2003), and the present results suggest that these responses may contribute to reward learning. Furthermore prediction errors activated the lateral prefrontal cortex during appetitive conditioning in the present study and in a study investigating causal learning (Fletcher et al. 2001). The cerebellum, which has primarily been implicated in coding aversive and motor prediction errors (e.g., Ploghaus et al. 2000), showed activations in the present study on reward-related learning as in a previous study on appetitive prediction errors (O'Doherty et al. 2003). Taken together, prediction error coding may constitute a basic form of brain functioning used throughout the brain in a wide variety of learning situations.

ACKNOWLEDGMENTS

We thank A. Dickinson and B. Seymour for helpful discussions and E. Featherstone for expert technical assistance.

Present address of J. P. O'Doherty: Div. of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125.

GRANTS

The study was supported by the Wellcome Trust, the Swiss National Science Foundation, and the Roche Research Foundation. R. J. Dolan is supported by a Wellcome Trust Programme Grant.

REFERENCES

- Amenta F, Barili P, Bronzetti E, Felici L, Mignini F, and Ricci A. Localization of dopamine receptor subtypes in systemic arteries. *Clin Exp Hypertens* 22: 277–288, 2000.
- Bao S, Chan VT, and Merzenich MM. Cortical remodelling induced by activity of ventral tegmental dopamine neurons. *Nature* 412: 79–83, 2001.
- Barto AG. Adaptive critics and the basal ganglia. In: *Models of Information Processing in the Basal Ganglia*, edited by Houk JC, Davis JL, and Beiser DG. Boston, MA: MIT Press, 1995, p. 215–232.
- Boynton GM, Engel SA, Glover GH, and Heeger DJ. Linear systems analysis of functional magnetic resonance imaging in human V1. *J Neurosci* 16: 4207–4221, 1996.
- Brembs B, Lorenzetti FD, Reyes FD, Baxter DA, and Byrne JH. Operant learning in Aplysia: neuronal correlates and mechanisms. *Science* 296: 1706–1709, 2002.
- Brett M, Anton J-L, Valabregue R, and Poline J-B. Region of interest analysis using an SPM toolbox (Abstract). Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2–6, 2002, Sendai, Japan. Available on CD-ROM Neuroimage 16, 2002.
- Dale AM and Buckner RL. Selective averaging of rapidly presented individual trials using fMRI. *Hum Brain Mapp* 5: 329–340, 1997.
- De Houwer J, Thomas S, and Baeyens F. Associative learning of likes and dislikes: a review of 25 years of research on human evaluative conditioning. *Psychol Bull* 127: 853–869, 2001.
- Dickinson A. Causal learning: an associative analysis. *Q J Exp Psychol B* 54: 3–25, 2001.
- Fletcher PC, Anderson JM, Shanks DR, Honey R, Carpenter TA, Donovan T, Papadakis N, and Bullmore ET. Responses of human frontal cortex to surprising events are predicted by formal associative learning theory. *Nat Neurosci* 4: 1043–1048, 2001.
- Friston KJ, Holmes AP, Worsley KJ, Poline JP, Frith CD, and Frackowiak RSJ. Statistical parametric maps in functional imaging: a general linear approach. *Hum Brain Mapp* 2: 189–210, 1994.
- Groves PM, Linder JC, and Young SJ. 5-hydroxydopamine-labeled dopaminergic axons: three-dimensional reconstructions of axons, synapses and postsynaptic targets in rat neostriatum. *Neuroscience* 58: 593–604, 1994.
- Hinich J, Lovibond PF, and Ter-Horst KM. Blocking in human electrodermal conditioning. *Q J Exp Psychol B* 48: 2–12, 1995.
- Hughes A, Thom S, Martin G, Redman D, Hasan S, and Sever P. The action of a dopamine (DA1) receptor agonist, fenoldopam in human vasculature in vivo and in vitro. *Br J Clin Pharmacol* 22: 535–540, 1986.
- Josephs O and Henson RN. Event-related functional magnetic resonance imaging: modelling, inference and optimization. *Philos Trans R Soc Lond B Biol Sci* 354: 1215–1228, 1999.
- Kamin LJ. Predictability, surprise, attention and conditioning. In: *Punishment and Aversive Behavior*, edited by Campbell BA and Church RM. New York: Appleton-Century-Crofts, 1969, p. 279–296.
- Knutson B, Fong GW, Bennett SM, Adams CM, and Hommer D. A region of mesial prefrontal cortex tracks monetarily rewarding outcomes: characterization with rapid event-related fMRI. *Neuroimage* 18: 263–272, 2003.
- Ljungberg T, Apicella P, and Schultz W. Responses of monkey dopamine neurons during learning of behavioral reactions. *J Neurophysiol* 67: 145–163, 1992.
- Logothetis NK, Pauls J, Augath M, Trinath T, and Oeltermann A. Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412: 150–157, 2001.
- Lynd-Balta E and Haber SN. The organization of midbrain projections to the ventral striatum in the primate. *Neuroscience* 59: 609–623, 1994.
- Mackintosh NJ. A theory of attention: variations in the associability of stimuli with reinforcement. *Psychol Rev* 82: 276–298, 1975.
- Martin I and Levey AB. Blocking observed in human eyelid conditioning. *Q J Exp Psychol B* 43: 233–256, 1991.
- McClure SM, Berns GS, and Montague PR. Temporal prediction errors in a passive learning task activate human striatum. *Neuron* 38: 339–346, 2003.
- McCoy AN, Crowley JC, Haghighian G, Dean HL, and Platt ML. Saccade reward signals in posterior cingulate cortex. *Neuron* 40: 1031–1040, 2003.
- Montague PR, Dayan P, and Sejnowski TJ. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J Neurosci* 116: 1936–1947, 1996.
- O'Doherty JP, Dayan P, Friston K, Critchley H, and Dolan RJ. Temporal difference models and reward-related learning in the human brain. *Neuron* 38: 329–337, 2003.
- O'Doherty JP, Dayan P, Schultz J, Deichmann R, Friston K, and Dolan RJ. Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* 304: 452–454, 2004.
- O'Doherty JP, Deichmann R, Critchley HD, and Dolan RJ. Neural responses during anticipation of a primary taste reward. *Neuron* 33: 815–826, 2002.

- Pasupathy A and Miller EK.** Different time courses of learning-related activity in the prefrontal cortex and striatum. *Nature* 433: 873–876, 2005.
- Pavlov IP.** *Conditional Reflexes*. London: Oxford UP, 1927.
- Pearce JM and Hall G.** A model of Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol Rev* 87: 532–552, 1980.
- Ploghaus A, Tracey I, Clare S, Gati JS, Rawlins JN, and Matthews PM.** Learning about pain: the neural substrate of the prediction error for aversive events. *Proc Natl Acad Sci USA* 97: 9281–9286, 2000.
- Rescorla RA and Wagner AR.** A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: *Classical Conditioning. II. Current Research and Theory*, edited by Black AH and Prokasy WF. New York: Appleton Century Crofts, 1972, p. 64–99.
- Reynolds JNJ, Hyland BI, and Wickens JR.** A cellular mechanism of reward-related learning. *Nature* 413: 67–70, 2001.
- Schoenbaum G, Setlow B, Saddoris MP, and Gallagher M.** Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdala. *Neuron* 39: 855–867, 2003.
- Schultz W.** Multiple reward signals in the brain. *Nat Rev Neurosci* 1: 199–207, 2000.
- Schultz W, Dayan P, and Montague PR.** A neural substrate of prediction and reward. *Science* 275: 1593–1599, 1997.
- Schultz W and Dickinson A.** Neuronal coding of prediction errors. *Annu Rev Neurosci* 23: 473–500, 2000.
- Sutton RS and Barto AG.** Toward a modern theory of adaptive networks: expectation and prediction. *Psychol Rev* 88: 135–170, 1981.
- Thorndike EL.** *Animal Intelligence: Experimental Studies*. New York: Macmillan, 1911.
- Thorpe SJ, Rolls ET, and Maddison S.** The orbitofrontal cortex: neuronal activity in the behaving monkey. *Exp Brain Res* 49: 93–115, 1983.
- Tremblay L and Schultz W.** Reward-related neuronal activity during go-nogo task performance in primate orbitofrontal cortex. *J Neurophysiol* 83: 1864–1876, 2000a.
- Tremblay L and Schultz W.** Modifications of reward expectation-related neuronal activity during learning in primate orbitofrontal cortex. *J Neurophysiol* 83: 1877–1885, 2000b.
- Waelti P, Dickinson A, and Schultz W.** Dopamine responses comply with basic assumptions of formal learning theory. *Nature* 412: 43–48, 2001.
- Wickens JR, Begg AJ, and Arbuthnott GW.** Dopamine reverses the depression of rat corticostriatal synapses which normally follows high-frequency stimulation of cortex in vitro. *Neuroscience* 70: 1–5, 1996.
- Williams SM and Goldman-Rakic PS.** Widespread origin of the primate mesofrontal dopamine system. *Cereb Cortex* 8: 321–345, 1998.